Bayesian Inference for Vertex-Series-Parallel Partial Orders

Abstract

Partial orders are a natural model for the social hierarchies that may constrain "queue-like" rank data. However, the computational cost of counting the LE's of a general partial order on a ground set with more than a few tens of elements is prohibitive. Vertex-series-parallel partial orders (VSP's) are a subclass of partial orders which admit rapid counting and represent the sorts of relations we expect to see in a social hierarchy. However, no Bayesian analysis of VSP's has been given to date. We construct a marginally consistent family of priors over VSP's with a parameter controlling the prior distribution over VSP depth. The distribution function is given in closed form. We extend an existing observation model for queue-like rank data to represent noise in our list-data and carry out Bayesian inference on rank lists from "Royal Acta" data. Model comparison shows our model is a better fit to the data and also compares favorably with a Plackett-Luce mixture.

1 INTRODUCTION

Rank data are analysed in a wide range of areas, including decision support, medical research, chemistry. Ranking methods can be classified into two categories - total order ranking and partial order ranking.

Total order models seek a ranking of the elements of the ground set (labels of the actors in a queue) that is "central" to the rank-lists in the data. These models are suitable when we can assume that an order relation exists between every pair of actors. Relevant models include the Mallow's model (Mallows [1957]) and the Plackett-Luce model (Plackett [1975]& Luce [1959]) and related mixture models. However, the real-world relations we are looking to recover may be weaker than a total order: relations between pairs of actors are not simply uncertain, they don't exist. We expect this for precedence relations that define some social hierarchies.

If we want to learn relations between actors by observing their behavior, then the elements of the model we fit should correspond to elements of reality: relations between actors must be represented by a *partial order*. A partial order $h = \{[n], \prec_h\}$ is a (possibly imcomplete) set of binary order relations $\prec_h$ over a "ground set" of actors with labels $[n] = \{1, \ldots, n\}$. We will observe the actors in queues constrained by a social hierarchy h, which is unknown. If we see enough queue realisations we can identify the hierarchy. In this setting the queue is just a *linear extension* (LE) of h, that is, a permutation of actors in $[n]$ that doesn't put an actor ahead of someone of higher precedence.

This leads to an analysis in which the partial order h is the parameter of interest, the data are LE's and the likelihood depends on the number of LE's of h. However, counting LE's is an NP-complete task [Brightwell and Winkler, 1991]. Work to date in this setting either restricts the class of partial orders [Mannila and Meek, 2000, Gionis et al., 2006, Mannila, 2008] to orders which admit fast counting or works with orders of manageable size [Beerenwinkel et al., 2007, Sakoparnig and Beerenwinkel, 2012, Nicholls and Muir Watt, 2011, Nicholls et al., 2022]. This approach does not scale well with n. We follow Mannila and Meek [2000] and work with *vertex-series-parallel* partial orders (VSP's). These orders are a sub-class of partial orders which can be formed by repeated *series* and *parallel* operations on smaller VSP's. They include *bucket orders* [1] as a special case. Valdes et al. [1979] represent VSP's using binary decomposition trees (BDT's). These support counting in a time linear in n [Wells, 1971] and scale to VSP's with hundreds of actors.

VSP's are a well characterised combinatorial class [Valdes et al., 1979, Wells, 1971]. However, work on fitting VSP's to

[1] Actors are grouped in buckets - every actor is ordered with respect to actors in other groups, and any pair of actors in the same group are incomparable.

Submitted to the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023). **To be used for reviewing only**.

data is limited. Mannila and Meek [2000] learn VSP's from LE's by adapting a greedy search over VSP's. However, there is no Bayesian inference. This is important because no prior probability distributions over VSP's existed. Mannila [2008] gave Bayesian inference for bucket orders, a subclass of VSP, and Sakoparnig and Beerenwinkel [2012] for partial orders, a super-set that doesn't scale.

Contributions. This is the first Bayesian inference for VSP's from LE's and presents some useful new priors and likelihoods. VSP's are equivalently a class of Directed Acyclic Graph (DAG); care is need when we specify priors over objects of this sort. Our prior (Sec. 2) is marginally consistent. This property is needed for the model to make sense in our setting. Our prior accurately represents the information available: it is non-informative with respect to VSP depth. Our prior probability mass function is available in a simple closed form. Our new observation model (Sec. 4), allowing for errors in queue-like data, is an improvement on previous models. We give a latent-space MCMC algorithm in C which targets the VSP posterior. Finally, our reconstruction of relations between witnesses appearing in Royal Acta (Sec. 5.2) is new. Historians are interested in these relations, but it wasn't possible to reconstruct them all till now as the partial orders were too big to count [Nicholls et al., 2022].

1.1 BACKGROUND

A partial order $h = \{V, \prec_h\}$ assigns a binary order relation[2] $\prec_h$ over a "ground set" of actors V. In our setting the actor labels are $V = [n]$ where $[n] = \{1, 2, ..., n\}$ or some subset. Partial orders on $[n]$ are in one-to-one correspondence with transitively-closed directed acyclic graphs (DAG's) with nodes $[n]$ and edges $E = \{\langle i, j\rangle \in [n] \times [n] : j \prec_h i\}$. Denote by $\mathcal{H}_{[n]}$ the set of all partial orders on actor labels $[n]$. Two actors $i, j \in [n]$ are *incomparable* $i \|_h j$, if neither $i \prec_h j$ nor $i \succ_h j$. Let $\mathcal{P}_{[n]}$ be the set of all permutations of $[n]$. A linear extension $l_h \in \mathcal{P}_{[n]}$ is a permutation of actors in $[n]$ that does not violate partial order h. See Fig. 1 for an example partial order[3] and its LE's. We denote the set of all LE's for partial order h as $\mathcal{L}[h]$. A *sub-order* $h[o] = (o, \prec_h)$ of a partial order $h \in \mathcal{H}_{[n]}$ restricts h to a subset $o \subseteq [n]$, $o = \{o_1, ..., o_m\}$ say: all order relations in h are inherited by $h[o]$ so the DAG representation $(o, E[o])$ has edges $E[o] = \{e \in E : e \in o \times o\}$; all edges involving elements in $[n] \setminus o$ are removed and all others remain. A *chain* of h is a sub-order $h[o]$ that is also a total order. The *length* of a chain is the number of nodes $|o|$ in the sub-order. The length of the longest chain is called the *depth*, $D(h)$ say, of the partial order, with $1 \leq D(h) \leq n$ for $h \in \mathcal{H}_{[n]}$.

[2]The binary relation $\prec_h$ is both irreflexive (the relation $i \prec_h i$ does not exist) and transitive (if $i \prec_h j$ and $j \prec_h k$, then $i \prec_h k$), where $i, j, k \in [n]$ and $i \neq j \neq k$.

[3]In this article, we visualise a partial order via its transitive reduction - this omits all edges implied by transitivity and is unique.

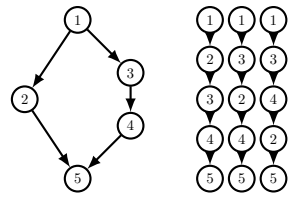

Figure 1: A partial order with 5 actors which is a VSP, v_0 say (left), and its three LE's (right).

The *vertex-series-parallel partial orders (VSP)* on $[n]$ are a large class of partial orders $\mathcal{V}_{[n]} \subset \mathcal{H}_{[n]}$ formed by repeated *series* $\otimes$ and *parallel* $\oplus$ operations. For partial orders h_1 and h_2, let $V(h_1)$ and $V(h_2)$ represent the ground sets of actors for h_1 and h_2 respectively (which we assume are disjoint).

- A *series partial order*, $h = h_1 \otimes h_2$, is the union of all relations in h_1 and h_2, with additional relations $i \succ j$ if $i \in V(h_1)$ and $j \in V(h_2)$.
- A *parallel partial order*, $h = h_1 \oplus h_2$, is the union of all relations in h_1 and h_2 with incomparability $i \| j$ if $i \in V(h_1)$ and $j \in V(h_2)$.

The set of VSP's is defined recursively as follows: if $|V(h)| = 1$ then h is a VSP; if h_1 and h_2 are VSP's then $h_1 \otimes h_2$ and $h_1 \oplus h_2$ are VSP's.

The partial order in Fig. 1 is a VSP, say v_0, can be constructed using the series and parallel operations in Fig. 2.

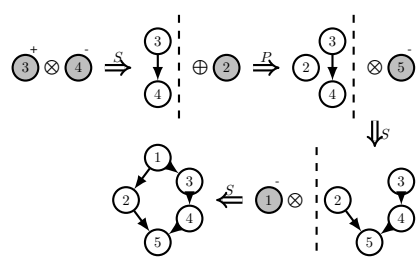

Figure 2: One possible construction procedure for VSP v_0.

A VSP on n actors can be parameterised as a Binary Decomposition Tree (BDT) Valdes et al. [1979] - a binary tree $t \in \mathcal{T}_{[n]}$ with n leaves. Edges are directed from the root to the leaves and nodes have additional attributes given below. Let $\mathcal{F}$ and $\mathcal{A}$ be the index sets for the n leaves and $n-1$ internal nodes respectively, with $\mathcal{F} \cup \mathcal{A} = [2n-1]$. Each leaf node index corresponds to a unique actor in the VSP. For $i \in \mathcal{F}$ let $F_i(t) \in [n]$ give the actor represented by leaf node i. Internal nodes $i \in \mathcal{A}$ are S nodes if the subtrees rooted by their child nodes are merged in series, and otherwise they are P nodes and the subtrees are merged in parallel. Internal nodes with an S label have an additional attribute indicating which of its child nodes is the "upper child": the subtree of this child node (indicated by a '+' in Fig. 3) is stacked above the subtree rooted by the other child node (indicated by a '-'). As an example, the VSP v_0 in Fig. 1 can be represented by the BDT in Fig. 3. Let $S(t) \in [n-1]$ be the number of S-nodes in tree t. A tree t with edge set $E(t)$ is

written $t = (F(t), E(t), L(t))$. Here $L(t) = \{L_i\}_{i \in \mathcal{A}}$ with $L_i(t) = (j, j')$ indicating that internal node i is an S-node with child nodes j, j' and the subtree rooted by j is stacked above that rooted by j', and $L_i(t) = \emptyset$ if i is a P-node. The map from a BDT to the VSP $v : \mathcal{T}_{[n]} \to \mathcal{V}_{[n]}$ is not bijective: For a VSP on n actors $v \in \mathcal{V}_{[n]}$, there exist many BDT's $t \in \mathcal{T}_{[n]}$. Let $t(v) = \{t \in \mathcal{T}_{[n]} : v(t) = v\}$.

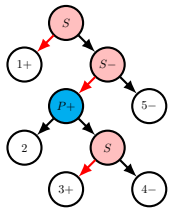

Figure 3: A BDT t_0 representing v_0 in Fig. 1, so that $v(t_0) = v_0$. Red edges and '+' signs indicate the upper child.

Brightwell and Winkler [1991] show that counting the number of LE's of a partial order is a #P-complete problem. However, the subclass of VSP partial orders admit fast counting. Wells [1971] give

$$|\mathcal{L}(h_1 \otimes h_2)| = |\mathcal{L}(h_1)||\mathcal{L}(h_2)| \tag{1}$$

$$|\mathcal{L}(h_1 \oplus h_2)| = |\mathcal{L}(h_1)||\mathcal{L}(h_2)| \binom{|V(h_1)| + |V(h_2)|}{|V(h_1)|} \tag{2}$$

where $|V(h_1)|$ and $|V(h_2)|$ give the number of actors in h_1 and h_2. This may be evaluated recursively in $O(n)$ steps.

2 VSP PRIOR

In this section we give a marginally consistent prior $\pi_{\mathcal{V}_{[n]}}(v|q)$ over VSP's on actors in $[n]$, controling the distribution over VSP-depth. We begin by defining a prior probability distribution $\pi_{\mathcal{T}_{[n]}}(t|q)$ over BDT's $t \in \mathcal{T}_{[n]}$.

Our prior on $\mathcal{T}_{[n]}$ has a uniform distribution over trees $([2n-1], E(t))$ with distinguishable leaves. Internal nodes are labelled S with probability q and otherwise P. We choose an "upper child" for each S node at random from its two child nodes, so

$$\pi_{\mathcal{T}_{[n]}}(t|q) = \frac{1}{|\mathcal{T}_{[n]}|} \left(\frac{q}{2}\right)^{S(t)} (1-q)^{n-S(t)-1}, \tag{3}$$

where $S(t)$ is a number of S-nodes and $|\mathcal{T}_{[n]}| = (2n-3)!! \equiv (2n-3)(2n-5)...3 \cdot 1$ is the number of binary tree topologies with n distinguishable leaves. The types of the $n-1$ internal nodes are independent with a factor $2^{-S(t)}$ for the stacking order of the children of S-nodes.

The prior on VSP's $t \in \mathcal{V}_{[n]}$ is then

$$\pi_{\mathcal{V}_{[n]}}(v|q) = \sum_{t \in t(v)} \pi_{\mathcal{T}_{[n]}}(t|q) \tag{4}$$

This choice, based on a uniform distribution over tree topologies, represents prior knowledge well.

A family of priors like $\pi_{\mathcal{T}_{[n]}}(t|q)$ or $\pi_{\mathcal{V}_{[n]}}(v|q)$, $n \geq 1$ is *marginally consistent* (also known as *projective*) if every marginal of every distribution in the family is also in the family. Marginal consistency need not hold if we simply write down a distribution over the trees or VSP's on any given set of actors, as we have done.

Definition 1 (Marginal consistency) *Let $\mathcal{O}_{[n]} = \{o \subseteq [n] : |o| > 0\}$ be the set of all subsets of $[n]$ with at least one element. The family of VSP priors $\pi_{\mathcal{V}_o}(v|q)$, $o \in \mathcal{O}_{[n]}$, $n \geq 1$ is marginally consistent if, for all $n \geq 1$ and all $o, \tilde{o} \in \mathcal{O}_{[n]}$ with $o \subseteq \tilde{o}$, distributions in the family satisfy*

$$\pi_{\mathcal{V}_o}(w|q) = \sum_{\substack{v \in \mathcal{V}_{\tilde{o}} \\ v[o]=w}} \pi_{\mathcal{V}_{\tilde{o}}}(v|q) \quad \textit{for all } w \in \mathcal{V}_o. \tag{5}$$

If marginal consistency holds for all q then it holds for marginals $\pi_o(w)$ by taking expectations over q in (5).

We want prior beliefs about relations between actors to be marginally consistent: if we have *full knowledge* of order relations between a pair of actors, then these should not change if we subsequently learn about relations with a third actor. Adding an actor to a partial order would in general create new order relations through transitivity *unless those relations were already present*. VSP's have this property. Relations between actors in a VSP $v \in \mathcal{V}_{[n]}$ are determined by the type of their "Most Recent Common Ancestor" (MRCA) in any BDT $t \in t(v)$ representing v. For example the MRCA of actors 2 and 4 in the tree t_0 in Fig. 3 is the blue P-node, so $2\|_{v_0}4$ in the VSP v_0 in Fig. 1. Adding or removing a leaf in the BDT doesn't change relations between other actors because it doesn't change the types of their MRCA's. This property leads to marginal consistency of trees and VSP's.

The following Theorem is our first result: we give a closed form expression for the prior for a VSP and show that the family of priors is marginally consistent. For $v \in \mathcal{V}_{[n]}$ let $t \in t(v)$ be some tree representing v. Partition the internal nodes $\mathcal{A}$ of t into *S-clusters* $C_k^{(S)}$, $k = 1, ..., K_S$ and *P-clusters* $C_k^{(P)}$, $k = 1, ..., K_P$. An S-cluster is a maximal set of internal nodes of type S which are connected by edges in $E(t)$. The P-clusters are defined similarly. Adjacent nodes not belonging to the same cluster must be in clusters of different type. We will see (in Lemma 1 below) that two BDT's representing the same VSP have the same numbers of S and P clusters, with the same sizes.

Theorem 1 *The family, $\pi_{\mathcal{V}_o}(v|q)$, $o \in \mathcal{O}_{[n]}$ $n \geq 1$, of VSP priors is marginally consistent. The probability distribution over VSP's $v \in \mathcal{V}_{[n]}$ (in (4)) is*

$$\pi_{\mathcal{V}_{[n]}}(v|q) = \pi_{\mathcal{T}_{[n]}}(t|q) \prod_{k=1}^{K_P} (2|C_k^{(P)}|-1)!! \prod_{k'=1}^{K^S} \mathcal{C}_{|C_{k'}^{(S)}|} \tag{6}$$

where t may be taken to be any tree $t \in t(v)$ with P- and S-clusters defined above, $\pi_{\mathcal{T}_{[n]}}(t|q)$ is given in (3) and

$$\mathcal{C}_s = \frac{1}{s+1}\binom{2s}{s}, \quad s \geq 0 \tag{7}$$

is the s'th Catalan number [Stanley and Weisstein, 2002].

Proof 1 (Theorem 1) *In Proposition 2 we show that the family of tree-priors $\pi_{\mathcal{T}_{[n]}}(t|q)$, $o \in \mathcal{O}_{[n]}$, $n \geq 1$ is marginally consistent. This result is used in Proposition 3 to show VSP's are marginally consistent - the first part of Theorem 1. In Proposition 4 we show that all trees $t \in t(v)$ have equal values of $\pi_{\mathcal{T}_{[n]}}(t|q)$, so that $\pi_{\mathcal{V}_{[n]}}(v|q) = |t(v)|\pi_{\mathcal{T}_{[n]}}(t|q)$ for any $t \in t(v)$. Finally, in Proposition 5, we give a formula for $|t(v)|$. The Catalan numbers enter because $\mathcal{C}_{s-1}$ gives the number of BDT's representing a total order on s elements (see proof Proposition 5). This gives (6) and completes the proof of Theorem 1.*

Our VSP-prior gives good control over partial order depth (Fig.I.1) . We can choose the prior distribution over q so that the marginal distribution $\pi_{\mathcal{V}_{[n]}}(v)$ has a reasonably flat distribution over the depth $D(v)$ of the VSP-partial order v. This ensures the prior is non-informative with respect to partial-order depth, a property of a social hierarchy on actors which is of particular interest.

The remainder of this section is given over to establishing the propositions referred to in the proof of Theorem 1. We first give a process realising $t \sim \pi_{\mathcal{T}_{[n]}}(t|q)$ in which leaves are added to the tree one at a time. This construction appears in Valdes [1978] but without the random element.

Definition 2 (Leaf Insertion and deletion) *If $t' \in \mathcal{T}_{[n-1]}$, $t' = (F(t'), E(t'), L(t'))$, is a tree on actors $(i_1, ..., i_{n-1})$ with $\mathcal{F}' \cup \mathcal{A}' = [2n-3]$ then the* leaf-insertion *operation $t = t' \triangleleft (e, i_n)$ at edge $e = \langle e_1, e_2 \rangle$, $e \in E(t')$, gives a tree $t = (F(t), E(t), L(t))$ with two new nodes $j' = 2n-2$ and $j = 2n-1$, leaves $\mathcal{F} = \mathcal{F}' \cup \{j'\}$, internal nodes $\mathcal{A} = \mathcal{A}' \cup \{j\}$, leaf-to-actor map $F_{\mathcal{F}}(t) = F_{\mathcal{F}}(t')$ and $F_{j'}(t') = i_n$, edge set*

$$E(t) = E(t') \setminus \{e\} \cup \{\langle j, j' \rangle, \langle e_1, j \rangle, \langle j, e_2 \rangle\}$$

and $L(t) = L(t') \cup L_j$ where $L_j = (j', e_2), (e_2, j')$ or $\emptyset$ with probabilities $q/2, q/2$ and $1-q$ respectively. The leaf deletion operation $t' = t \triangleright i_n$ reverses this operation, pruning the leaf for actor i_n (and removing its parent node).

Definition 3 (Generative model for BDT's) *Let $(i_1, \ldots, i_n) \in \mathcal{P}_{[n]}$ be the actor list taken in any order. Simulate $t \sim \pi_{\mathcal{T}_{[n]}}(t|q)$ as follows:*

1. *Set $\mathcal{F} = \{0, 1\}$, $\mathcal{A} = \emptyset$, $F_1 = 1, E = \{\langle 0, 1 \rangle\}, L = \emptyset$ and $t_{(1)} = (F, E, L)$ (a single-edge tree);*
2. *For $k = 2 : n$, (add the actors one at a time)*
   (a) *choose an edge $e \sim \mathcal{U}\{E(t_{(k-1)})\}$ at random;*
   (b) *set $t_{(k)} = t_{(k-1)} \triangleleft (e, i_k)$;*
3. *$E(t_{(n)})$ contains an edge $e = \langle 0, e_2 \rangle$. Return the BDT $t = (F(t_{(n)}), E(t_{(n)}) \setminus \{e\}, L(t_{(n)}))$ with leaf labels $\mathcal{F} \leftarrow \mathcal{F} \setminus \{0\}$.*

If we run this generative model we get a random tree distributed according to $\pi_{\mathcal{T}_{[n]}}$.

Proposition 1 (Prior Probability Distribution over $\mathcal{T}_{[n]}$) *The probability distribution over BDT's determined by the model in Definition 3 is given by (3). (See A.1 for proof.)*

We now define sub-trees of BDT's. At the end of step k in the tree-generation process in Definition 3 the "current tree" is $t_{(k)} \in \mathcal{T}_o$ with $o = (i_1, ..., i_k)$ and at the end of step $k' > k$ it is $t_{(k')} \in \mathcal{T}_{\tilde{o}}$ with $\tilde{o} = (i_1, ..., i_k, i_{k+1}, ..., i_{k'})$. If, for $o, \tilde{o} \in \mathcal{O}_{[n]}$ with $o \subseteq \tilde{o}$, we fix $\tau \in \mathcal{T}_o$ and $t \in \mathcal{T}_{\tilde{o}}$ then the conditional probability

$$\pi_{\mathcal{T}_{\tilde{o}}|\mathcal{T}_o}(t|\tau, q) = \Pr(t_{(k')} = t | t_{(k)} = \tau, q)$$

is the probability to realise $t_{(k')} = t$ when $t_{(k)} = \tau$.

Definition 4 (Sub-trees and containing trees) *Tree τ is a sub-tree of t (and t contains τ) if $\pi_{\mathcal{T}_{\tilde{o}}|\mathcal{T}_o}(t|\tau, q) > 0$. Let*

$$\mathcal{T}_{\tilde{o}}(\tau) = \{t \in \mathcal{T}_{\tilde{o}} : \pi_{\mathcal{T}_{\tilde{o}}|\mathcal{T}_o}(t|\tau, q) > 0\}$$

give the set of trees in $\mathcal{T}_{\tilde{o}}$ containing a given tree $\tau \in \mathcal{T}_o$.

If t contains τ then t can be realised from τ by a sequence of edge insertions $\triangleleft$ and τ can be recovered from t removing the actors in $\tilde{o} \setminus o$ using the pruning operator $\triangleright$.

The family of prior distributions over trees $\pi_{\mathcal{T}_o}(\tau|q)$, $o \in \mathcal{O}_{[n]}$, $n \geq 1$ is marginally consistent if, for all $n \geq 1$ and all $o, \tilde{o} \in \mathcal{O}_{[n]}$ with $o \subseteq \tilde{o}$, distributions in the family satisfy

$$\pi_{\mathcal{T}_o}(\tau|q) = \sum_{t \in \mathcal{T}_{\tilde{o}}(\tau)} \pi_{\mathcal{T}_{\tilde{o}}}(t|q) \quad \text{for all } \tau \in \mathcal{T}_o\,. \tag{8}$$

Proposition 2 *The probability distribution over BDT's given in (3) is marginally consistent. (See A.2 for proof.)*

Proposition 3 *The probability distribution over VSP's given in (4) is marginally consistent. (See A.3 for proof.)*

This concludes the first part of Theorem 1. We now prove the second part. Valdes [1978] showed that *Multi-Decomposition Trees* (MDT's) are one-to-one with VSP's.

These trees are obtained by collapsing edges which connect internal nodes of the same S/P-type in the BDT, as in Fig. A.1. A formal definition is given in A.4 . It is then straightforward to count the number of BDT's that reduce to a given MDT.

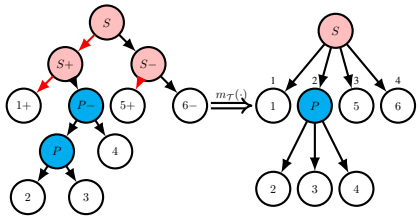

Figure 4: An example BDT t_1 (left) and its corresponding MDT $m_1 = m_{\mathcal{T}}(t_1)$ (right). The child-nodes of any S-node in the MDT are numbered to give the order in which their subtrees are stacked.

For $m \in \mathcal{M}_{[n]}$ the mapping $v = v(m)$ determines a unique VSP $v \in \mathcal{V}_{[n]}$, as relations between any pair of actors in the VSP are given by the type of their MRCA in m. Every VSP has a MDT, as it has a BDT which may be collapsed to a MDT. Let $m_{\mathcal{V}}(v) = \{m \in \mathcal{M}_{[n]} : v(m) = v\}$ be the set of MDT's representing $v \in \mathcal{V}_{[n]}$.

Lemma 1 *The map $m_{\mathcal{V}} : \mathcal{V}_{[n]} \to \mathcal{M}_{[n]}$ is one to one so we may write $m = m_{\mathcal{V}}(v)$. (This result is discussed in Valdes et al. [1979]. See Valdes [1978] for proof.)*

We next observe that all BDT's representing the same VSP have equal prior probabilities (they collapse to the same MDT and that fixes $S(t)$). This makes it easy to do the sum in (4) as the summand is constant.

Proposition 4 (Probability Distribution over VSP's) *The prior probability for a VSP with n nodes is*

$$\pi_{\mathcal{V}_{[n]}}(v|q) = |t(v)|\pi_{\mathcal{T}_{[n]}}(t|q),$$

for any tree $t \in t(v)$. (See A.5 for proof.)

Finally, we count trees in $t(v)$ and this gives us the closed form we seek. This seems to be new.

Proposition 5 *Let $t \in t(v)$ be an arbitrary BDT of a VSP $v \in \mathcal{V}_{[n]}$ with P- and S-clusters defined as in Theorem 1. The number of BDT's of v is*

$$|t(v)| = \prod_{k=1}^{K_P} (|2C_k^{(P)}| - 1)!! \prod_{k'=1}^{K^S} \mathcal{C}_{|C_{k'}^{(S)}|} \tag{9}$$

with $\mathcal{C}_s,\ s \geq 0$ given in (7). (See A.6 for proof.)

This completes the proof of Theorem 1. Since VSP's and MDT's are one-to one, we can parameterise our posterior in Sec. 4 in terms of MDT's. We do this in Corollary A.1 in A.7. This gives $S(t),\ t \in t_{\mathcal{M}}(m)$ directly so we can evaluate $\pi_{V_{[n]}}(v|q)$ without constructing a BDT t.

3 BI-DIRECTIONAL QUEUE-JUMPING OBSERVATION MODEL

Our data is a collection of N lists. For $j \in [N]$ let $o_j \subseteq [n],\ o_j = \{o_1, ..., o_{n_j}\}$ be the actors present when the j'th ranking list was observed and let $y_j \in \mathcal{P}_{o_j},\ y_j = (y_{j,1}, ..., y_{j,n_j})$ be the observed list, just an ordered version of o_j. Let $y = (y_1, ..., y_N)$ be the list of lists. The 'queue-based' observation model given in Nicholls and Muir Watt [2011] models list data as a realisation of a random queue constrained to put higher status individuals before those of lower status. Let $x \in \mathcal{P}_{\tilde{o}}$ be a single list ranking actors $\tilde{o} \subseteq [n]$ and $v \in \mathcal{V}_{[n]}$ a VSP constraining actors, and having a suborder $v[\tilde{o}]$ over the actors in $\tilde{o}$. The noise-free observation model derived from the queue process has $x \sim \mathcal{U}(\mathcal{L}[v[\tilde{o}]])$, so x is uniform on the LE's of the suborder $v[\tilde{o}]$ constraining the actors $\tilde{o}$ actually present.

It is unlikely the observations are "error free". In a "queue-jumping" model an error parameter $p \in [0, 1]$ gives the probability the "next" person in the queue jumped the queue. The queue can be built from the bottom up or top down, and so Nicholls et al. [2022] give two versions of this model, one in which actors jump up the queue (QJ-U, see B.1) and one in which actors jumping down the queue (QJ-D).

We would like to have a queue-based model with a tractable likelihood in which displacement can occur in both directions. The 'bi-directional queue-jumping' model QJ-B combines the 'jumping-up' and 'jumping-down' models. It is not simply a mixture of the two distributions, as it allows displacement in both directions within a single realisation. The cost of evaluating a QJ-B likelihood is exponential in n. However, for the application of interest there is a subset of actors who are known a priori to appear as a group. The size of this subset is manageable ($n \simeq 20$). Although QJ-B cannot be evaluated for a general partial order (counting LE's is prohibitive) it is fine for a VSP. Separate modelling of this subset is well-motivated.

Like QJ-U, QJ-B ranks by repeated selection. A list $x \in \mathcal{P}_{[n]}$ is built up from both ends. Let $z \in \{0, 1\}^{n-1}$ with $z_k \sim Bern(\phi)$. Here $z_k = 0$ indicates the k'th actor to be added to the list was placed bottom-up in the QJ-D model and $z_k = 1$ indicates they were placed top-down in the QJ-U model. If $U_0 = 0$ then $U_k = U_{k-1} + z_k$ gives the number of places filled from the top after the k'th actor has arrived, so the k'th actor was placed into position U_k in x. Similarly, if $D_0 = n+1$ then $D_k = D_{k-1} - (1 - z_k)$ tracks places filled from the bottom and gives the index in x when $z_k = 0$. Let $i_k = z_k U_k + (1 - z_k) D_k$ give the position in x into which the k'th actor was added. For example if $z = (1, 0, 1, 1, 0, 0, 1)$ then $(i_1, ..., i_n) = (1, 7, 2, 3, 6, 5, 4)$ and after step 3 the list was $(x_1, x_2, \emptyset, \emptyset, \emptyset, \emptyset, x_7)$ with $x_3, ..., x_6$ not yet realised.

Definition 5 (Bi-Directional Queue-Jumping Model) *Let $L_T(v) = |\mathcal{L}[v]|$ be the number of LE's of VSP $v \in \mathcal{V}_{[n]}$*

and for $i \in [n]$ let $T_i(v) = |\{l \in \mathcal{L}[v] : l_1 = i\}|$ give the number of LE's with actor i at the top. Let $B_i(v) = |\{l \in \mathcal{L}[v] : l_n = i\}|$ give the number of LE's with actor i at the bottom. If $z \in \{0,1\}$ is given then $i_k = i_k(z), k = 1, ..., n$ is given above. The observation model for QJ-B for a list $x \in \mathcal{P}_{[n]}$ given z is

$$Q_{bi}(x|z,v,p,\phi) = \prod_{k=1}^{n-1}[\phi \mathbb{1}_{\{z_k=0\}} Q_{bi}(x_{i_k}|x_{i_{1:k-1}}, z_k, v, p) + (1-\phi)\mathbb{1}_{\{z_k=1\}} Q_{bi}(x_{i_k}|x_{i_{1:k-1}}, z_k, v, p)],$$

where

$$Q_{bi}(x_{i_k}|x_{i_{1:k-1}}, z_k = 0, v, p) = \frac{p}{n-k+1} + (1-p)\frac{T_{x_{i_k}}(v[x_{[n]\setminus\{i_1,...,i_{k-1}\}}])}{L_T(v[x_{[n]\setminus\{i_1,...,i_{k-1}\}}])}$$

$$Q_{bi}(x_{i_k}|x_{i_1:k-1}, z_k = 1, v, p) = \frac{p}{n-k+1} + (1-p)\frac{B_{x_{i_k}}(v[x_{[n]\setminus\{i_1,...,i_{k-1}\}}])}{L_T(v[x_{[n]\setminus\{i_1,...,i_{k-1}\}}])},$$

and marginally,

$$Q_{bi}(x|v,p,\phi) = \sum_{z\in\{0,1\}^n} Q_{bi}(x|z,v,p,\phi)p(z|\phi) \tag{10}$$

where $p(z|\phi) = \phi^{\sum_i z_i}(1-\phi)^{n-\sum_i z_i}$.

We give a generative model realising $x \sim Q_{bi}$ in B.2. This distribution reduces to Q_{up}/QJ-U in B.1 when $\phi = 1$ (and Q_{down}/QJ-D when $\phi = 0$). We use this nesting to investigate whether QJ-U or QJ-D or QJ-B fits the data better. This is of interest in our application as different error types correspond to obvious physical mechanisms (downwards displacement may be "arrived late" and upwards displacement may be "my friend the King is present"). When $p = 0$ this is the noise free model for every $\phi \in [0,1]$, so ϕ is not identifiable in the noise free setting.

Counting LE's of a VSP (evaluating $L_T(v)$ etc) is $O(n)$ so the computational complexity for naive evaluation of Q_{bi} using (10) is $O(n^2 2^n)$. We used a recursion (Algorithm B.3) of computational complexity $O(n2^n)$. This avoids repeated evaluation of LE-counts for the same suborders and (by Proposition 7 in B.3) evaluates Q_{bi}.

4 SUMMARISING THE VSP POSTERIOR

Bayesian inference is straightforward in principal given an explicit prior distribution over VSP's and an observation model $Q = Q_{up}$ or $Q = Q_{bi}$ for our N ranking-lists. In fact we can either represent a VSP as an MDT (since the mapping is one-to-one) or carry out Bayesian inference on the latent space of BDT's $t \in \mathcal{T}_{[n]}$ and use the fact that they marginalise to MDT's. We present both posteriors. Let $\psi = (p, \phi)$ for QJ-B and $\psi = p$ for QJ-U.

The posterior for the BDT $t \in \mathcal{T}_{[n]}$ is

$$\pi_{\mathcal{T}_{[n]}}(t,q,\psi|y) \propto \pi_{\mathcal{T}_{[n]}}(t|q)\pi(q,\psi)Q(y|v(t),\psi) \tag{11}$$

The posterior distribution for the VSP $v \in \mathcal{V}_{[n]}$ is

$$\pi_{\mathcal{V}_{[n]}}(v,q,\psi|y) \propto \pi_{\mathcal{V}_{[n]}}(v|q), \pi(q,\psi)Q(y|v,\psi). \tag{12}$$

where in practice we use the equivalent MDT posterior with prior given in Corollary A.1 for VSP's in (12).

Proposition 6 (Posterior Marginals) *Sampling the BDT posterior $(t,q,\psi) \sim \pi_{\mathcal{T}_{[n]}}(\cdot|y)$ gives samples $(v(t),q,\psi) \sim \pi_{\mathcal{V}_{[n]}}(\cdot|y)$ from the VSP posterior (see A.8 for proof).*

5 APPLICATIONS

5.1 DATA AND ANALYSES

We analyse a dataset accessed through a database made for 'The Charters of William II and Henry I' project by the late Professor Richard Sharpe and Dr Nicholas Karn [Sharpe et al., 2014]. These data collect witness lists from legal documents from England and Wales in the eleventh and twelfth century. Witness lists respect a rigid social hierarchy: higher status individuals come ahead of lower status individuals in the lists. Fig. D.1 in Sec. D is an example list.

We represent the hierarchy on actors $[n]$ as a partial order which is a VSP $v \in \mathcal{V}_{[n]}$ and model a list as the outcome of a queuing process. Actors $o \subset [n]$ are present. They form a (virtual) queue, exchanging positions at random, subject to respecting the VSP. It is straightforward to show that the equilibrated queue is a uniform draw from the LE's of the suborder $v[o]$. Occasional departures from the rules of placement are allowed by the error model.

Lists are witnessed by people from all walks of life and we have their titles. These include "others" (actors who lack titles). Historians are interested in social hierarchies and how they change over time. For illustration we reconstruct hierarchies in three snapshots: the years 1080-84, 1126-30 and 1134-38. The last two cover periods shortly before and after Stephen becoming King, a time of great change. The 5-year intervals are short enough for any changes in the hierarchy to be slight [Nicholls et al., 2022]. For ease of visualisation we present results for individuals appearing in at least 5 lists (5LPA data) here and results on all actors (1LPA data) in Sec. D.1. We fit VSP/QJ-U to all data and fit VSP/QJ-B to 2 of the 3 5LPA data sets (not 1134-38, as QJ-B has runtime growing exponentially with the length of the longest list). However, relations between bishops in 1134-38 are of particular interest so we present VSP/QJ-B results for this subgroup. Table D.1 in Sec. D summarises the data in the different experiments.

The prior for error probability p and for q (probability for an S-node) is given in Fig. 7. All fitting is done using MCMC

in the BDT representation, Algorithm C.1. For any given model we draw samples $t^{(k)}, p^{(k)}, \phi^{(k)} \sim \pi_{\mathcal{T}_{[n]}}(\cdot|y)$ for $k = 1, ..., K$ and set $v^{(k)} = v(t^{(k)})$ per Proposition 6. Example MCMC traces are given in the supplement with Effective Sample Size (ESS) values (Supplement D.2). Sampled VSP's are summarised using consensus VSP's: $V^{con}(\epsilon)$ includes order relation/edge $\langle i, j \rangle$ if the relation appears more than ϵK times in the MCMC output. We color edges black if they are in $V^{con}(\epsilon)$ at $\epsilon = 0.5$ but not $\epsilon = 0.9$ and red if they are supported at $\epsilon = 0.9$. We plot transitive reductions, so many red edges from the top to the bottom of the DAG are omitted for ease of viewing.

In Sec. 5.2, we apply QJ-U the QJ-B models to the 5LPA data and make a model comparison using Bayes factors. Consensus orders for the 1LPA data are given in Supplement D.1. We compare these models with bucket order models and a Plackett-Luce mixture in Supplement Sec. F and Sec. E respectively. We report computing time measurements for counting LE's for the standard partial order model and the VSP is compared empirically in Supplement Sec. G.

5.2 RESULTS

Consensus orders for actors color-coded by their professions are shown in Fig. 5 and 6. For both QJ-U and QJ-B models, we observe three clear social hierarchies: King $\succ$ Queen $\succ$ Duke appear at the top, in that order (when they are in the 5LPA data, in 1180-84 and 1134-38); then archbishop/prince $\succ$ bishops; the remaining professions (earl, count, chancellor, other) are ranked lower than bishops in a relatively complex hierarchy.

Some of this is common sense. However, the web of strongly attested relations between earls and others in 1134-38 is new. There is clear evidence for hierarchies within professions. The bishop-only analysis in 1134-38 is similar to but does not quite match the bishop subgraph of the full analysis for the same period. Information is shared across lists so removing actors changes estimated order relations between those that remain (and can reduce the number of lists). However, the bishops appear as a group, there are few non-bishops "between" bishops in lists, so this effect is slight. We can attach names to nodes: for example, the top three bishops in 34-38 are Henry, de Blois, Bishop of Wincester $\succ$ Roger, Bishop of Salisbury $\succ$ Alexander, Bishop of Lincoln.

The reconstructed status hierarchy by QJ-B is simpler and deeper than QJ-U (Fig 6). The data must contain a small number of errors in both directions. A uni-directional model must fit a shallower VSP as it accommodates errors in the "wrong" direction by removing order relations in the reconstructed VSP. The consensus order for the 1134-38(b) data is simpler but qualitatively similar to the reconstructed under the QJ-U model.

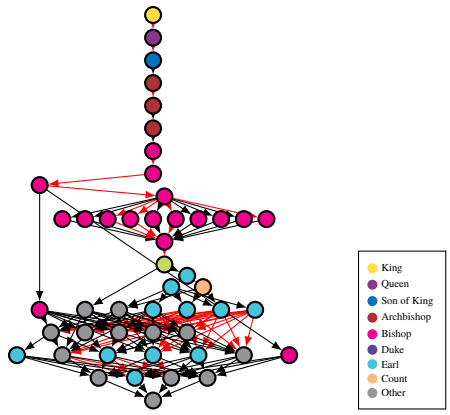

Figure 5: VSP/QJ-U model. Consensus order for 1134-38 5LPA data. Significant/strong order relations are indicated by black/red edges respectively.

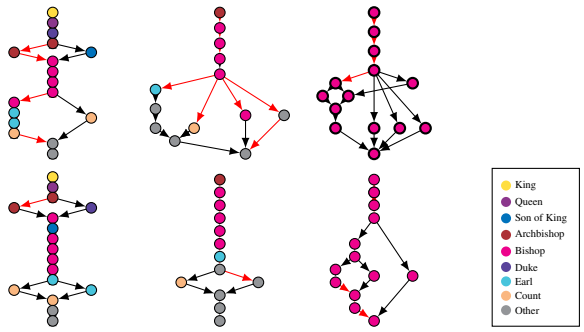

Figure 6: VSP/QJ-U (top row) and VSP/QJ-B (bottom row). Consensus orders for 1180-84, 1126-30 and 1134-38 (bishops) (left to right columns) 5LPA data. Legend as Fig. 5.

We summarise "professions" within VSP's by averaging ranks. Given a partial order $v \in \mathcal{V}_{[n]}$, we rank an actor $i \in [n]$ by the number of actors above them, $\text{rank}_i(v) = 1 + |\{\langle e_1, e_2 \rangle \in E(v) : e_2 = i\}|$, and take as our summary the average rank of actors in the profession. The posterior mean average ranks are given in Table D.6 and D.8 and they match our remarks on the consensus orders.

Prior and posterior distributions for q (probability for serial node, $P(S)$), error probability p and queue jumping up probability ϕ for the three periods are shown in Fig. 7. The p-posteriors are weighted toward smaller values and overlap, though errors are low in 1126-30 and higher in 1180-84 indicating greater respect for the rules of precedence in 1126-30 than in 1180-84. The prior and posterior depth distribution are shown in Fig. D.11 and D.14 . The prior depth distributions are fairly flat so any depth-structure in the posterior comes from the data. The probability for serial operation (q) closely relates to the depth of order relation. For example, in 1180-84 a relatively high q for QJ-U is associated with relatively high depth VSP's (given the maximum value) with a maximum depth of 14 over 17 actors. In contrast, the posterior probabilities for S and P operators are almost equal

for 1134-38 and relative to the maximum depth we get a shallower hierarchy: the posterior mean for depth is about 23 over 49 actors in Figure D.11.

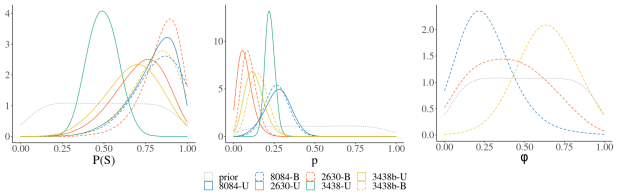

Figure 7: Posterior distributions for $q = P(S)$ (left), error probability p (middle) and queue jumping up probability ϕ (right) for the time periods 1180-1184 (blue), 1126-1130 (red), 1134-1138 (green) and 1134-1138(b) (yellow) from both the VSP/QJ-U (solid) and VSP/QJ-B (dashed) models. The prior is represented in grey in all figures.

The QJ-B allows both error directions. The error-direction tends to be up if $\phi > 1/2$ and otherwise down so QJ-U is favored during 1134-38(b), while the 1080-84 data aligns more with QJ-D. In contrast to QJ-U, more serial operations (high q) supports deeper order relation reconstruction in Fig. 6. However, the p-posteriors between the two models are similar. At small p the displacement direction controlled by ϕ is hard to measure and not identifiable at $p = 0$ so the ϕ-distributions are correspondingly broad.

Preference shifts from downwards to bidirectional to upwards displacement error models over the period 1080-1140. We justify this below using Bayes factors. In summary, QJ-D is slightly favored over QJ-B ($D > B$) in 1080-84 while in 1126-30 models QJ-D and QJ-B are equally good ($D \approx B$). Both are clearly favored over QJ-U in these periods ($D, B \gg U$). In 34-38(b) we have $U \approx B$ and $U, B \gg D$. This coincides to the elpd comparison in Table E.1. As an example, in order to compute the Bayes factor $B_{U,B}$ for up against bi-directional displacement we consider

$$B_{U,B} = \frac{p(y|\phi = 1)}{p(y|\phi \in (0,1))} = \lim_{\delta \to 0} \frac{\pi(\phi > 1 - \delta|y)}{\pi(\phi > 1 - \delta)}.$$

From Fig. 7, $B_{U,B}$ is close to 0 in periods 1180-84 and 1126-30 as the posterior density is well below the prior density at $\phi \to 1$, providing strong support for QJ-B over QJ-U. In 1134-38(b), we see $B_{U,B} \approx 1$, as the curves meet as $\phi \to 1$ so there is no clear signal from the data. The other comparisons may be justified similarly.

We make comparisons with a Plackett-Luce mixture model (E) and with Bucket orders (F). With Bayes factors around 2 or 3, Bucket orders are equal or slight preferred over VSP in the QJ-B model. VSP are clearly preferred in the QJ-U fit (some very large Bayes factors in favor of VSP).

Our list data are incomplete, in the sense that the membership in list $i = 1, ..., N$ is o_i not $[n]$ and N-values in Table D.1 are not much larger than the number of actors n.

In order to measure the reliability of our reconstructions we take parameters sampled from the posterior (the last sampled state $v^{(K)}, p^{(K)}, q^{(K)}, \phi^{(K)}$) and generate synthetic data with the same list-membership and length structures as the real data. We apply our methods to each synthetic data set in Sec. D.2.1. The ROC curves Fig. D.12 and D.15 for consensus orders $V^{con}(\epsilon)$ show the proportion of inferred false-positive and true-positive relations increasing with decreasing ϵ from $(0, 0)$ at $\epsilon = 1$ (the consensus order is empty) to $(1, 1)$ at $\epsilon = 0$ (complete graph). For each simulated data set there is ϵ giving high true-positive and low false-positive reconstructed relation fractions.

6 DISCUSSION AND CONCLUSION

Our work was motivated by the need to fit relatively large partial orders (up to 200 nodes) to noisy linear-extension data. We saw that, for data on this scale, counting extensions using the VSP-tree representation is much faster than the current state-of-art on partial orders, enabling our methods to scale. We gave a new consistent and closed form prior distribution over VSP's with a parameter q controlling VSP depth, and a new observation model QJ-B for noisy LE's which generalises QJ-U [Nicholls and Muir Watt, 2011]. We fit the new model to some of the smaller data sets and the old model to all data sets. Neither of these analyses would be possible without the VSP-setup. The data support the new observation model in our application. Our $elpd_{waic}$-based model comparisons also clearly favor QJ-U and QJ-B over a Plackett-Luce mixture. Although we could fit the large data sets, visualising consensus partial orders proved challenging and this is an area we need to improve.

We gave MCMC algorithms targeting the posterior for VSP's in both the latent-space (BDT) parameterisation and the integrated MDT parameterisation. We found the BDT-MCMC adequate, though it would be good to make an efficiency comparison with MDT-MCMC, which we expect to be more efficient. BDT updates which don't change the VSP are fast so our simple approach will be close. For code see https://github.com/JessieJ315/Bayesian-Inference-for-Vertex-Series-Parallel-Partial-Orders.git.

In future work we would like to compare our fit with the recently-proposed contextual repeated selection (CRS) model (Seshadri et al. [2020] and Ragain and Ugander [2018]). This is a rich class of models for rank-order data. The elements of the model are not essentially physical, in the sense that a VSP represents a social hierarchy relation by relation. Also, CRS models do not encode transitivity. It is easy to show VSP models cannot be represented as CRS models with "cliques" of size two (a system of equations linking the CRS and VSP distributions over extensions has no solutions). However, CRS models may fit the data well, and a comparison would be revealing.

References

Niko Beerenwinkel, Nicholas Eriksson, and Bernd Sturmfels. Conjunctive bayesian networks. *Bernoulli*, pages 893–909, 2007.

Graham Brightwell and Peter Winkler. Counting linear extensions. *Order*, 8(3):225–242, 1991.

Francois Caron and Arnaud Doucet. Efficient bayesian inference for generalized bradley–terry models. *Journal of Computational and Graphical Statistics*, 21(1):174–196, 2012.

Aristides Gionis, Heikki Mannila, Kai Puolamäki, and Antti Ukkonen. Algorithms for discovering bucket orders from data. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 561–566, 2006.

Kustaa Kangas, Teemu Hankala, Teppo Mikael Niinimäki, and Mikko Koivisto. Counting linear extensions of sparse posets. In *IJCAI*, pages 603–609, 2016.

R Duncan Luce. On the possible psychophysical laws. *Psychological review*, 66(2):81, 1959.

Colin L Mallows. Non-null ranking models. i. *Biometrika*, 44(1/2):114–130, 1957.

Heikki Mannila. Finding total and partial orders from data for seriation. In *International Conference on Discovery Science*, pages 16–25. Springer, 2008.

Heikki Mannila and Christopher Meek. Global partial orders from sequential data. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 161–168, 2000.

Cristina Mollica and Luca Tardella. Bayesian plackett-luce mixture models for partially ranked data. *Psychometrika*, 82(2):442–458, 2017. ISSN 0033-3123. doi: 10.1007/s11336-016-9530-0.

Cristina Mollica and Luca Tardella. Plmix: An r package for modelling and clustering partially ranked data. *Journal of Statistical Computation and Simulation*, 90(5):925–959, 2020.

Geoff K Nicholls and Alexis Muir Watt. Partial order models for episcopal social status in 12th century england. *IWSM 2011*, page 437, 2011.

Geoff K. Nicholls, Jeong Eun Lee, Nicholas Karn, David Johnson, Rukuang Huang, and Alexis Muir-Watt. Bayesian inference for partial orders from random linear extensions: power relations from 12th century royal acta, 2022. URL https://arxiv.org/abs/2212.05524.

Robin L Plackett. The analysis of permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 24(2):193–202, 1975.

Stephen Ragain and Johan Ugander. Choosing to rank. *arXiv preprint arXiv:1809.05139*, 2018.

Thomas Sakoparnig and Niko Beerenwinkel. Efficient sampling for Bayesian inference of conjunctive Bayesian networks. *Bioinformatics*, 28(18):2318–2324, 07 2012. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts433. URL https://doi.org/10.1093/bioinformatics/bts433.

Arjun Seshadri, Stephen Ragain, and Johan Ugander. Learning rich rankings. *Advances in Neural Information Processing Systems*, 33:9435–9446, 2020.

R. Sharpe, D. Carpenter, H. Doherty, M. Hagger, and N. Karn. The charters of William II and Henry I. Online: Last accessed 27 October 2022, 2014.

Richard Stanley and Eric W. Weisstein. *Catalan Number*. https://mathworld.wolfram.com/CatalanNumber.html, 2002. MathWorld–A Wolfram Web Resource.

Jacobo Valdes. *Parsing Flowcharts and Series-Parallel Graphs*. PhD thesis, Stanford, CA, USA, 1978. AAI7905944.

Jacobo Valdes, Robert E Tarjan, and Eugene L Lawler. The recognition of series parallel digraphs. In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 1–12, 1979.

Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, 27:1413–1432, 2017.

Mark B Wells. Elements of combinatorial computing. 1971.